

Exploring How Pipe-Level Energy Metrics Relate to Hydraulic Parameters Using Multivariate Analysis

Saeed Hashemi¹, Yves R. Filion², Vanessa L. Speight³

¹ Water Design Specialist, CH2M Hill
Toronto, Ontario (Canada)

² Department of Civil Engineering, Queen's University
Kingston, Ontario, (Canada)

³ University of Sheffield, Sheffield, (UK)

¹ saeed.hashemi@ch2m.com, ² yves.filion@civil.queensu.ca, ³ v.speight@sheffield.ac.uk

ABSTRACT

PCA in this study builds upon a correlation matrix of 11 hydraulic parameters including pipe hydraulic data and pipe-level energy metrics by using Eigen values and Eigen vectors to compress the high-dimensional space of the data into a two-dimensional space. This makes the visualization of the data possible, including all hydraulic parameters simultaneously. The graphical presentation of the hydraulic parameters (mono-plots) and the data set (bi-plots) help which parameters can lead into inefficiency in pipes in the whole dataset.

The results show that the metric energy lost to friction in a pipe along with average unit headloss, average flow rate and proximity to major components have a high influence in distinguishing poorly performing pipes from the others. Average pressure and the metric for energy needed by the user for each pipe tend to track closely, despite a lower statistical importance than previous parameters. Diameter and pipe roughness tend to stand alone with poor representations on the two principal component axes.

Keywords: Energy Efficiency; Principal Components Analysis; Water Distribution Systems.

1 INTRODUCTION

Aging water infrastructures face water main breaks, major leaks and considerable headlosses regularly [1, 2]. Therefore, in order to stay functioning and cost effective, these infrastructures require continual monitoring and scheduled rehabilitation [3, 4]. Aging water mains from late 1800's in North American cities importune municipalities, while cities are almost always under budget to keep up with maintaining these systems [5]. It seems that there is a pressing need to create new knowledge in this regard, a missing link between pipe rehabilitation and energy dynamics of water mains. This new insight would be a new approach to better understanding pressure conduits and their energy behavioral character [6, 7, and 8].

Previous research has suggested the use of energy indicators to improve knowledge about the energy performance of pressure pipes in water distribution networks [6, 8, and 9]. These studies for the most part focus on energy dynamics of these systems at a network level. Even though a network approach is informative for decision makers to understand how their systems compare to ideal or desirable performance, there is also a need to understand individual pipe energy performance. Pipe-level energy indicators developed by Hashemi et al. [10] have corroborated previous research findings by Dziedzic

and Karney [8] that larger water mains closer to treatment plants tend to waste more energy. However, the results of the study by Hahsemi et al. [10] create new knowledge concerning the spatial and diurnal variation of energy efficiency in distribution systems.

In another study, Hashemi et al., [11] however, have performed correlation and regression analyses to relate their newly suggested pipe-level energy indicators to decision parameters in available to municipal managers. This study can be perceived as a step towards creating a knowledge applicable to other WDSs for the purpose of energy-based pipe rehabilitation. However, further assessment of the results shows that correlation analysis may not provide information on which hydraulic factors or energy metrics can better describe the variance or have a higher influence on the whole pipe data set on a broader sense [12, 13]. This is because correlation does not possess the capacity of considering a number of hydraulic parameters simultaneously.

The present study extends the previous work by applying principal components analysis (PCA) to visualize the relative importance of a number of hydraulic parameters in energy efficiency a large ensemble of pipes. This data sets comprises over 20,000 pipes water mains, selected from 17 systems in Ontario, Ohio and Kentucky [14]. The advantages of graphical presentation of the correlation matrix include: 1) an understanding of which hydraulic parameters and energy metrics are clustered together, 2) identification of similarities in describing the variance of the data, 3) a demonstration of which pipe parameters/energy metrics seem to behave differently from the other parameters, and 4) an estimate of the importance of each of the hydraulic parameters and energy metrics in describing the variance within a large ensemble of pipes.

2 Methods

2.1 Pipe-level energy metrics

Energy metrics that are going to be considered in the statistical analysis along with the hydraulic factors are listed and summarized and defined in Table 1. Other hydraulic factors are average daily pressure, average flow rate, average elevation of pipes, unit headloss, hydraulic proximity, pipe roughness and diameter. Hydraulic proximity to major components is also defined in Table 1, however, more details about this newly developed indicator and energy metrics can be found in another study by Hashemi et al. [10].

Table 1. Pipe- level energy metrics by Hashemi et al. [10]

| Energy Metric | Definition |
|-------------------------------|---------------------------------------------------------------------------------------|
| Gross Energy Efficiency (GEE) | $GEE = \frac{E_{\text{delivered}}}{E_{\text{supplied}}} \times 100\%$ |
| Net Energy Efficiency (NEE) | $NEE = \frac{E_{\text{delivered}}}{E_{\text{supplied}} - E_{\text{ds}}} \times 100\%$ |
| Energy Needed by User (ENU) | $ENU = \frac{E_{\text{delivered}}}{E_{\text{need}}} \times 100\%$ |
| Energy Lost to Friction | $ELTF = \frac{E_{\text{friction}}}{E_{\text{supplied}} - E_{\text{ds}}} \times 100\%$ |
| Hydraulic Proximity* | $Q \cdot H_s \text{ (m}^4/\text{s)}$ |

*Hydraulic proximity is not an energy metric like other items in this table.

2.2 Principal Components Analysis (PCA)

PCA is capable of visualizing the whole data set by compressing the transcription of the high dimension data set into lower dimensions, based on correlation matrix. This method finds new directions/dimensions on which the data set shows the highest variance/scatteredness [15]. These new directions are rotations of the previous data set with regard to previous directions/dimensions [16], which are hydraulic parameters and energy metrics, in this study. Therefore, each of the observations on the previous system of coordinates can be assigned a new score on the new system of coordinates, also known as Principal Components (PCs). Each of the new directions or PCs have a contribution in describing the variance of the data. The first few PCs describe most of the variance in the data set so that the rest of the PCs or dimensions are negligible. More important PCs or Eigen vectors correspond to larger Eigen values of the correlation matrix of the ranks of variables of interest [11]. In fact, by finding the Eigen values for correlation matrix of the variables of interest, directions on which correlations of variables are zero will be achieved, thus showing the most variance/scatteredness. This can distinguish pipes or groups of pipes from each other will be the point of interest in this study.

2.3 PCA Mono-plots and Bi-plots

Each of the components of the PCs are called a “loading”, implying the score of each variable of interest on the PCs. If the loading of a variable is high on a PC, it means that specific variable is well represented by that PC. Initial variables of interest with high scores on the first two PCs seem to be more influential in the data set. Plotting the first two PCs on a two dimensional plot, also known as a “mono-plot”, can represent the initial variables of interest and their influence on the new coordinate system (PCs). Variables that track closely together imply similarity in their effect on the data set, while those staying away from others display different influences. Moreover, the original observations in the data set can be presented in a “bi-plot” by transforming their original values into the new PC coordinates.

3 Application of Multivariate Statistical Analyses in Large WDSs

For the purpose of attaining robust results in statistical analysis, pipe data must be adaptable to different situations. 17 WDSs, possessing varying characteristics, have thus been selected from different areas in North America, such as states of Kentucky, Ohio, and the province of Ontario. This large ensemble of pipes includes over 20,000 pipes. The characteristics of these WDSs are summarized in Table 1. In order to obtain basic hydraulic results for mentioned WDSs, EPANET2.0 network models were used [17].

Table 1. Summary of characteristics of 17 North American WDSs

| Network | State/ Province | No. of Pipes | Pipes length (km) | No. of junctions | Difference in elevations (m) | No. of pumps | No. of Tanks | Average daily demand (MLD) | Average daily pressure (m) |
|----------------|------------------------------|-----------------------------|----------------------------------|-----------------------------|-------------------------------------------------|-------------------------|-------------------------|-----------------------------------------------|-----------------------------------------------|
| 1 | ON ₁ ^a | 12189 | 627 | 11177 | 50 | 31 | 10 | 69.07 | 44.86 |
| 2 | ON ₂ | 405 | 56 | 349 | 46 | 6 | 3 | 3.54 | 46.71 |
| 3 | KY ₁ ^b | 984 | 67 | 856 | 37 | 1 | 2 | 7.52 | 33.07 |
| 4 | KY ₂ | 1124 | 152 | 811 | 29 | 1 | 3 | 7.92 | 46.07 |
| 5 | KY ₃ | 366 | 91 | 271 | 43 | 5 | 3 | 15.19 | 41.76 |
| 6 | KY ₄ | 1156 | 260 | 959 | 75 | 2 | 4 | 5.65 | 48.02 |
| 7 | KY ₅ | 496 | 96 | 420 | 75 | 9 | 3 | 8.58 | 134 |
| 8 | KY ₆ | 644 | 123 | 543 | 96 | 2 | 3 | 6.19 | 60.2 |
| 9 | KY ₇ | 603 | 137 | 481 | 70 | 1 | 3 | 5.80 | 55.32 |
| 10 | KY ₈ | 1614 | 247 | 1325 | 135 | 4 | 5 | 9.32 | 54.15 |
| 11 | KY ₉ | 1270 | 972 | 1242 | 138 | 17 | 15 | 5.07 | 94 |
| 12 | KY ₁₀ | 1043 | 435 | 920 | 96 | 13 | 13 | 8.18 | 68 |
| 13 | KY ₁₁ | 846 | 464 | 802 | 248 | 21 | 28 | 6.61 | 97.11 |
| 14 | KY ₁₂ | 2426 | 655 | 2347 | 145 | 15 | 7 | 5.18 | 111 |
| 15 | KY ₁₃ | 940 | 155 | 778 | 95 | 4 | 5 | 8.92 | 50.78 |
| 16 | KY ₁₄ | 548 | 105 | 377 | 65 | 5 | 3 | 3.94 | 53.9 |
| 17 | OH ₁ ^c | 1183 | 166 | 956 | 100 | 15 | 4 | 10.13 | 57 |

a. ON= Ontario, b. KY= Kentucky, c. OH= Ohio

4 Results

The results for Eigen values of the correlation matrix, in Figure 2 also known as scree plot, show that the first two Eigen vectors/PCs in this case have the most contribution in describing the variance in the data (almost 65%). The first and second PCs according to this scree plot describe 47.3% and 16.9% of the variance in the data set, respectively. Hence, the first two PCs are selected to in order to transform, compress and visualize the pipe data set in the form a new two-dimensional space. Hydraulic parameters and energy metrics each have scores/loadings on the new PCs or in other words in the new coordinate system, summarized in Table 2. Mono-plots presented discussed in the next section are a visual expression of what is included in Table 2.

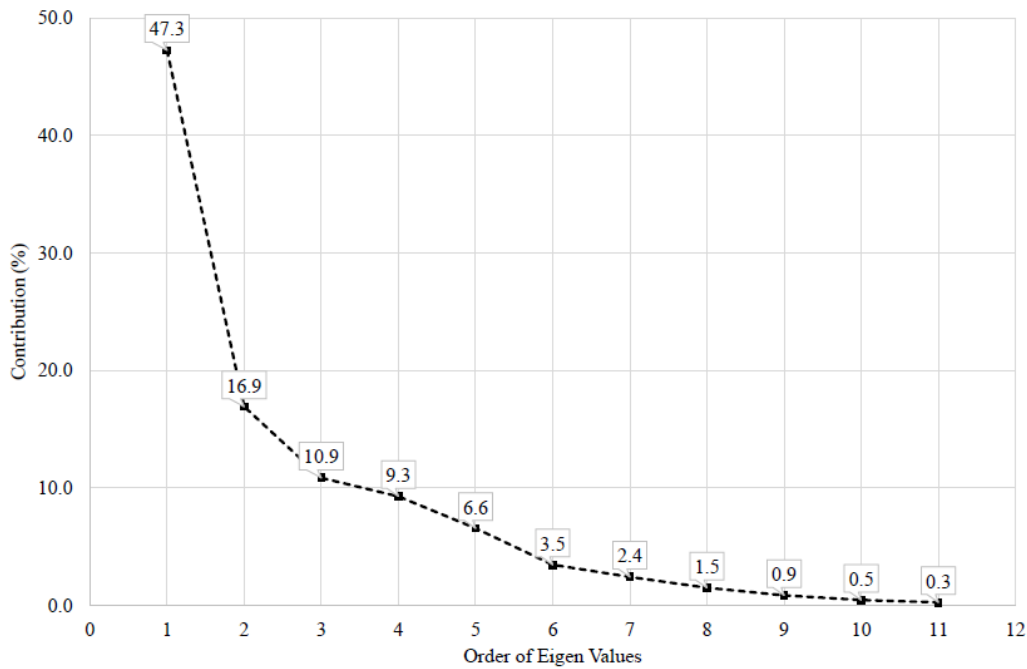


Figure 1. Scree plot- Contribution of Eigen Values/Principal components

Table 2. First two (most influential) PCs in the 17 systems

| | PC ₁ | PC ₂ |
|------------------|-----------------|-----------------|
| C | -0.04 | 0.12 |
| D | 0.23 | -0.06 |
| P | -0.06 | 0.59 |
| Q | 0.41 | 0.05 |
| headloss | 0.37 | 0.04 |
| Proximity | 0.40 | 0.05 |
| Elevation | 0.09 | -0.45 |
| GEE | -0.38 | -0.06 |
| NEE | -0.41 | -0.07 |
| ENU | -0.07 | 0.64 |
| ELTF | 0.39 | 0.07 |

4.1 Hierarchical Importance of Hydraulic Parameters

In the mono-plot presented in Figure 2, x-axis represents principal component 1 (PC₁) which shows the most variation in the pipe data set (47.3%). However, y-axis represents PC₂ describing the second most variation in the data (16.9%). This can also be a graphical presentation of Table 2.

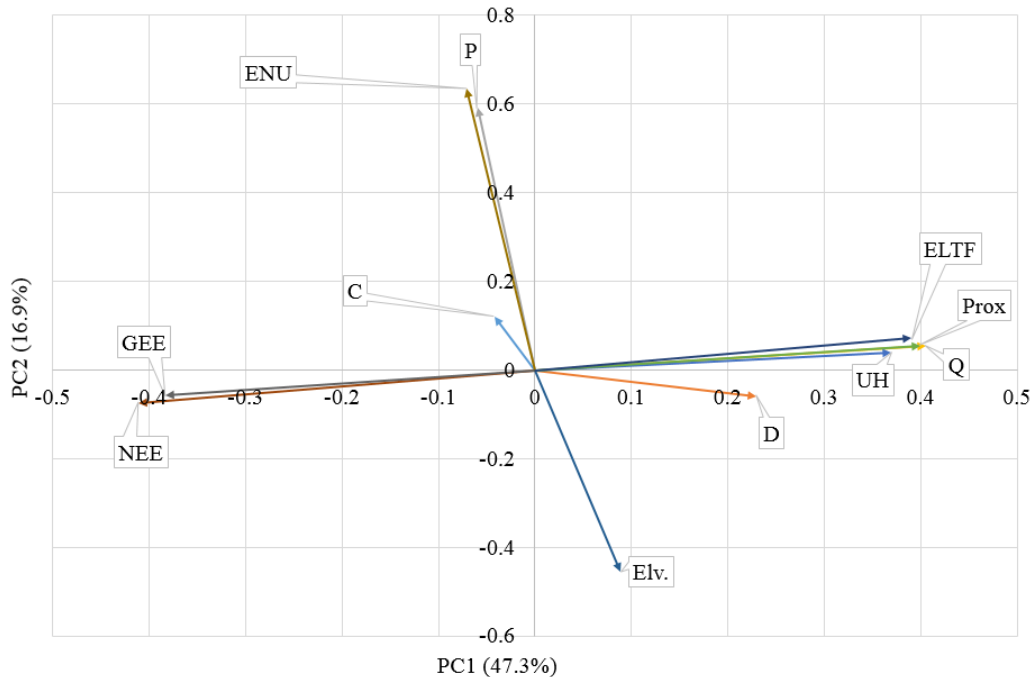


Figure 2. Mono-plot of energy metrics and hydraulic factors in 17 systems

According to Figure 2, *GEE* and *NEE* track closely, meaning that higher values for one results in higher values for the other too and vice versa. Also, based on the PC₁ values for these two parameters, they seem to be more influential in describing variance. On the other hand, *ELTF*, Ave Q, headloss and proximity cluster together. This not only means that they have similar effects on pipes but also, that high values of these parameters result in lower values of *GEE* and *NEE*. It is also noted that all parameters of *GEE*, *NEE*, *ELTF*, proximity, Ave Q and headloss are well represented with regard to PC₁, as they show relatively higher values of PC₁.

Moreover, along PC₂ axis, it seems that *ENU* and *P* have higher importance and cluster up. Therefore, it can be inferred that first, these two parameters are highly correlated to each other, and second they enjoy a higher importance compared to *D* and *C_{HW}*. However, *ENU* and *P* would have a lower priority compared to those with higher values of PC₁ (*ELTF*, *GEE*, *NEE*, headloss, etc.). This is mainly because PC₂ inherently describe less variance (almost 16.9%) compared to PC₁ (almost 47.3%).

C_{HW} and *D* do not seem to be close to any other parameters or each other, which means that they will not have an effect similar to any other parameters. Also, these parameters are not well represented on any other the two axes of PC₁ and PC₂, as a result of attaining low scores on these axes. Hence, they are not perceived influential on describing the variance of pipe data or decision making with regard to energy dynamics. The fact that elevation stands away from other parameters implies that it has a different effect on variance of the pipe dataset. Also from Figure 2, it can be inferred that higher elevations cause lower pressure, as the two parameters stand in the opposite directions of PC₂. Elevation in this study represents the average elevation of upstream and downstream nodes for each pipe.

Figure 3 shows the bi-plot of the pipes from the 17 systems. This figure shows the score of each pipes with regard to PC₁ and PC₂. The poor and ideal pipes in terms of energy metrics are then highlighted in red and green respectively. An important point in this discussion would be gleaned from stacking Figures 2 and 3 together, as they are based on the same axes/PCs (different scales). In fact, doing so would clarify what hydraulic parameters in Figure 2 would drive pipes towards green or red pipes in Figure 3.



Figure 3. Bi-plot of pipes on PCs- Ideal versus poor pipes

5 CONCLUSION

This study attempted to showcase the hierarchical importance of hydraulic parameters in explaining the energy efficiency of pipes. The results aimed for graphical presentation of correlations and a step further to analyze and monitor a number of factors simultaneously. In contrary to correlation results, PCA not only visualized the correlation of pairs of parameters, but also outlined the importance of each factor among all from a decision maker's perspective. PCA made this possible by compressing the transcription of the data with high dimensions. The results finally showed that hydraulic proximity, average flow and unit headloss can be more influential than average pressure, elevation roughness and diameter, in distinguishing the cohorts of poor pipes from efficient pipes.

6 ACKNOWLEDGEMENTS

The authors wish to thank the Natural Science and Engineering Research Council for its financial support of this research. The authors also thank CH2MHill Canada for their financial support to present this research study.

References

- [1] Dandy, G. C., and Engelhardt, M. (2001). Optimal scheduling of water pipe replacement using genetic algorithms. *Journal of Water Resources Planning and Management*, 127(4), 214-223.
- [2] Engelhardt, M. O., Skipworth, P. J., Savic, D. A., Saul, A. J., and Walters, G. A. (2000). Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. *Urban Water*, 2(2), 153-170.
- [3] Kleiner, Y., & Rajani, B. (2001). Comprehensive review of structural deterioration of water mains: statistical models. *Urban water*, 3(3), 131-150.
- [4] Lansey, K. E., Basnet, C., Mays, L. W., and Woodburn, J. (1992). Optimal Maintenance Scheduling for Water Distribution Systems. *Gordon and Breach Science Publishers*, U.K., 211-226.
- [5] Prosser, M. E., Speight, V. L., & Fillion, Y. R. (2013). Life-cycle energy analysis of performance-versus age-based pipe replacement schedules. *Journal of American Water Works Association*, 105, E721-E732.
- [6] Cabrera, E., Pardo, M. A., Cobacho, R., and Cabrera Jr, E. (2010). Energy audit of water networks. *Journal of Water Resources Planning and Management*, 136(6), 669-677.
- [7] Cabrera, E., Gómez, E., Cabrera Jr, E., Soriano, J., and Espert, V. (2014). Energy Assessment of Pressurized Water Systems. *Journal of Water Resources Planning and Management*, 141(8), 04014095: 1-12.
- [8] Dziejcz, R. M., and Karney, B. W. (2014). Water Distribution System Performance Metrics. *Procedia Engineering*, 89, 363-369.
- [9] Pelli, T., and Hitz, H. U. (2000). Energy indicators and savings in water supply. *Journal of the American Water Works Association*, 92(6), 55-62.
- [10] Hashemi, S., Fillion, Y. R., Speight, V. L. (2017a). Pipe-Level Energy Metrics to Assess the Impact of Infrastructure and Operational Decisions on Energy Use in Water Distribution Systems, *Journal of water resources planning and management* (Submitted).
- [11] Hashemi, S., Fillion, Y. R., Speight, V. L. (2017b). Examining the relationship between unit headloss and energy performance of water mains with statistical analyses across an ensemble of water distribution systems, *Journal of water resources planning and management* (Submitted).
- [12] Weiss, N. A., and Weiss, C. A. (2012). Introductory statistics. London: Pearson Education.
- [13] Edwards, A. (1984). An Introduction to Linear Regression and Correlation, Second Edition, W.H. Freeman and Company, New York.
- [14] Jolly, M. D., Lothes, A. D., Sebastian Bryson, L., and Ormsbee, L. (2013). Research database of water distribution system models. *Journal of Water Resources Planning and Management*, 140(4), 410-416.
- [15] Krzanowski, W. J. (1988). Principles of multivariate analysis, A user's perspective. *Clarendon Press*, Oxford.
- [16] Jolliffe, I.T. (1986). Principal Component Analysis. *Springer-Verlag*, New York, USA.
- [17] Rossman, L.A. (2000). EPANET2: User's Manual. *US Environmental Protection Agency*. Cincinnati, OH.